



Cross-dialectal vowel mapping and glide perception

Abram Clear & Anya Hogoboom*

Abstract. Formant transitions from a high front vowel to a non-high, non-front vowel mimic the formant signature of a canonical [j], resulting in the perception of an *acoustic glide* (Hogoboom 2020). We ask if listeners may still perceive a glide when canonical formant transitions are absent. We investigated the mapping of an Appalachian English (AE) monophthongal [aɪ] in hiatus sequences, monophthongal [aɪ.a]. If participants map this monophthongal [aɪ] to a high front position, they might perceive a glide that is not supported by the acoustic signal, which we call a *phantom glide*. Ninety-six participants (45 of which were native AE speakers) heard 30 different English words ending in [i], [ə], or monophthongal [aɪ] (i.e. *tree*, *coma*, *pie*) that had been suffixed with either [-a] or [-ja]. They were asked to identify which suffixed form they heard. Participants in both dialect groups sometimes perceived a glide that was truly absent from the speech stream. In these cases, participants mapped static formants in monophthongal [aɪ.a] stimuli to a diphthongal /aɪ/ with a high front endpoint, causing the perception of the necessary F1 fall and subsequent rise of a [j]. Using recent models of speech processing, which encode both social and acoustic representations of speech (e.g. Sumner et al. 2014), we discuss the mapping of monophthongal [aɪ] to a privileged diphthongal underlying form.

Keywords. glides; speech perception; Appalachian English; dialectal variation; vowel perception; exemplar theory; hiatus

1. Introduction.

1.1. GLIDES. Glides are semi-vowel consonants whose articulation is very similar to movement common in vowel-to-vowel transitions. They are described as “maximally similar to their vocalic environment” (Uffmann 2007:458). Past attempts to differentiate phonological glides from vowels in succession cite degrees of constriction, noting consonantal glides typically possess more frication and constriction than V.V sequences (Padgett 2008). Recent research has called this heuristic into question. For example, Burgdof and Tilson (2021) found high vowels to be significantly more constricted than glides. Using electromagnetic articulography, they also found glide productions to be more precise, having less variance in articulation than vowel productions. Regardless of fine-grained acoustic properties, glides and certain vowel sequences are perceptually confusable.

As the acoustic realization of glides mimics vowel transition, certain vowels in hiatus position (i.e. adjacent but parsed into separate syllables) can precipitate the acoustic waveform of and perception of a glide. Formant transitions from a high front vowel to a central or low back vowel mimic the formant signature of a canonical [j]. Though one lacks an intentional glide, forms like

*We thank members of the W&M Linguistics Labs, our poster audience at the 2021 meeting of the LSA, the Roy R. Charles Center for summer funding, and the moderators of the University of Kentucky Appalachian Studies listserv for participant recruitment. We extend a special thank you to Kate Harrigan for her invaluable assistance with experiment design. Thank you to Kim Love of K. R. Love Quantitative Consulting and Collaboration for statistical consulting. Authors: Abram Clear, William & Mary (awclear@email.wm.edu) & Anya Hogoboom, William & Mary (ahogoboom@wm.edu).

[ki.ə] and [ki.jə] are confusable to English speakers. Following Hogoboom (2020), we call these unintentional glides incurred from vowel transition *acoustic glides*.

This line of research partially investigates if we can perceptually disambiguate acoustic glides from phonologically present glides. Cross-linguistically, glide formation, /CV₁.V₂/ → [CGV₂], is a commonly employed method of hiatus resolution (Casali 2011). Homorganic glide epenthesis/formation has been argued to be less phonologically costly than the epenthesis of a new consonant segment (Uffmann 2007). Hogoboom (2020) argues phonological glides and high front to central or low back vowel hiatus sequences are confusable even for speakers of languages that contrast [i.a] and [i.ja] forms (i.e. Mandarin, Korean, etc.).

1.2. DIALECTAL VARIATION AND SPEECH PERCEPTION. This work uses the term “mapping” to refer to a complex set of processes that could be components of either generative or exemplar models of speech perception. Exemplar-based models (e.g. Goldinger 1996; Johnson 1997) allow for the storing of non-phonetic social factors and nuanced speaker information alongside mental representations. They differ from generative models in their fundamental building blocks, preferring word-size or sound-size exemplars versus abstracted phonemes. “Mapping” may be considered synonymous with “matching” of an exemplar at either the word or sound level. If attempting to fit our findings within a generative model, “mapping” may be equivalent to “abstracting to a prototype” or phoneme. These processes cannot be fully disambiguated as exemplar-based models can, on the surface, “exhibit abstraction behavior” due to exemplar storage (Johnson 2005: 301). We now turn to recent studies of the language systems’ processing of natural variation across dialects and individual speakers. Generally, these studies attempt to disambiguate the presence of one or two underlying representations for the perception dialectal variants.

Studying the processing of r-ful (ie. bak[ə]) and r-less (ie. bak[ə]) surface forms in populations of NYC English and General American (GA) English, Sumner and Samuel (2009) argued for the presence of different phonological representations for different dialectal forms. Overt speakers of New York City (NYC) English produced r-less forms but used both r-less and r-ful forms in perception, storing representations for both variants. The authors made a distinction between bidialectal speakers (Overt NYC speakers) and *fluent listeners* (Covert NYC speakers). Bidialectal speakers map NYC r-less and GA r-ful surface variants to separate lexical representations; fluent listeners easily encode surface variants as realizations of a single, r-ful underlying form. Beyond evidence for dialect specific encodings, Sumner and Samuel found a general benefit for unmarked variants (r-ful forms in their study), which supported previous findings allocating more perceptual weight and facilitative capacity to prestigious forms (McLennan et al. 2003; Ranbom & Connine 2007). In a later study, Sumner and Kataoka (2013) found a memory benefit for the same r-less form produced by a speaker of a prestigious dialect, non-rhotic British English.

Investigating the perception of variation in the language system of Basque-Spanish bilinguals, Samuel and Larraza (2015) argued for a dual-mapping route to a single exemplar, resulting in a quasi-allophonic representation. Teaching Basque-Spanish bilinguals novel words with a consistent form of either /t_ʂ/ or /tʃ/, they controlled the encoding of single representations into the lexical system of their participants. No participants encountered /t_ʂ/-/tʃ/ variation for any novel word. Samuel and Larraza tested the speakers’ acceptability judgements of “mispronounced” variants, substituting /t_ʂ/ for /tʃ/ and vice versa. The “mispronounced” variants were accepted as real Basque words with an error rate of ~30%. For purposes of lexical access, [t_ʂ] and [tʃ] are treated as essentially allophonic variants for Basque-Spanish bilinguals. This

underspecified model of dual-mapping to a single exemplar mechanically parallels the proposed language system of a *fluent listener* (Sumner & Samuel 2009).

Studying American English vowel perception, Fridland and Kendall (2011) concluded that “processing is affected by both what you say and what others around you say” (792). They presented participants from the U.S. South, North, and West with a forced choice vowel identification task. Participants were presented with a word containing a synthesized vowel at one of seven steps between a sample speaker’s natural endpoints of /e/ and /ɛ/. Participants were asked to identify the word as its /e/ or /ɛ/ form (ie. *bait* or *bet*). The authors also collected production data from a subset of their participants via a reading passage and word list. Most important for this research, Southern speakers’ productions were influenced by the Southern Vowel Shift; they produced the mid-front vowels, /e/ and /ɛ/ much closer together than their Northern and Western counterparts. Reflecting a perceptual system influenced by their own habits of production, Southerners were found to maintain /e/ perception along the continuum of synthesized vowels toward the /ɛ/ target longer than non-Southerners.

1.3. GLIDE PERCEPTION AND PERCEPTUAL MAPPING. Hogoboom (2020) established the confusability of acoustic glides and phonological glides. The rise of F1 and fall of F2 necessary to transition from [i] to [a], [o], or [ə] can incur the perception of an acoustic glide. We ask, could the *perception* of a high front starting position in a V.V sequence incur the reporting of a glide that is fully absent from the speech signal. Rather than present through formant mimicry or acoustic byproducts, we investigate the perception of glides that lack any acoustic foundation. We call these consonants that may be perceived via perceptual mapping and top-down processes of lexical access *phantom glides*. We hypothesize that phantom glides may be perceived by accessing underlying representation of a high front vowel (be that a sound/word-level exemplar or phoneme) when presented with a non-high vowel’s surface form. For example, American English speakers produce [e] as an inherent diphthong [eɪ]. They hear [eɪ] far more frequently than [e] and have no social impetus to privilege the [e] form. Therefore, they may map [e] to a higher and fronter underlying /eɪ/. In an [e.a] sequence, such a mapping, /eɪ.a/, could induce the perception of a rise in F1 and fall in F2. Such formant transitions are exactly what enables the perception of acoustic glides, and, in the [e.a] case, we might expect would incur the perception of a phantom glide.

This research pushes the question of perceptual mapping further. We ask if speakers can map an initial low front vowel in a hiatus sequence to a high front underlying form. This high front mapping could lead to the perception of formant transition in relatively static low vowel to low vowel hiatus sequences, which could incur the secondary effect of phantom glide perception. We use the monophthongal form of /aɪ/, which is discussed in detail below, to ascertain the existence of phantom glides. Due to the possible encoding of /aɪ/ as a prestige form over monophthongal /aɪ/ or the quasi-allophonic positioning of monophthongal [aɪ] in perceptual relation to diphthongal /aɪ/, listeners have the potential to map monophthongal [aɪ] to a diphthongal underlying form. Mapping to a diphthongal underlying form (with a high front end position) could incur the perception of phantom glides in certain hiatus contexts.

1.4. APPALACHIAN ENGLISH. Appalachian English (AE) is a dialect spoken throughout the southern portion of the Appalachian mountain range, including western North Carolina and Virginia, southern West Virginia, and Eastern Kentucky and Tennessee (Wolfram 1977). It is important to recognize this speaker population is not monolithic; some research has supported the differentiation of Appalachian English into separate Appalachian Englishes to better capture a multiplicity of linguistic and social identities within the Appalachian mountain range (Hazen 2020).

Most important to this research, Appalachian English speakers produce /aɪ/ as a monophthong with a low front target region (Reed 2014). Monophthongization of the /aɪ/ diphthong has been observed in speakers of broadly Southern American English (e.g., Bernstein 2006; Labov, Ash, & Boberg 2005; Greene 2010). However, it is far more common in southern Appalachian English productions, appearing in pre-voiceless consonant contexts as well as in pre-voiced and word-final environments (Thomas 2001).

Studies in perceptual dialectology (folk linguistics), have shown Southern speech features to be the most regionally marked and salient to non-linguists in the United States (Clopper & Pisoni 2004, 2007; Niedzielski & Preston 2003; Preston 1989, 1993). As Appalachia is stigmatized, most Appalachians have a high degree of linguistic awareness; they are conscious of how their language is negatively perceived by non-Appalachians (Greene 2010; Reed 2014, 2016). While possessing covert prestige within the region, Appalachian English features can index negative stereotypes of unintelligence, closed-mindedness, and depravity. Best summarized by Williams (2002), the negative stereotype faced by Appalachians is one of “the profligate hillbilly, amusing but often also threatening, defined by a deviance and aberration, a victim of cultural and economic deprivation attributable to mountain geography” (17).

Such prejudice encourages assimilation to more “standard,” prestigious linguistic forms, which may be produced by covert, diasporic Appalachians through dialectal code-switching. Reed (2018) has investigated the effects of *rootedness*, which he defines as one’s attachment to place, on monophthongal [aɪ] production. Reed (2018) found nearly categorical change in one speaker’s monophthongal [aɪ] production after she left Appalachia. The diphthongal, mainstream production of [aɪ] was “more fitting for her current social identity” and likely helped evade negative connotations of Appalachian identity (Reed 2018). Conversely, some native Appalachians retain Appalachian English speech forms in new social contexts for a sense of connection to their linguistic and cultural roots (Hazen & Hamilton 2008).

1.5. SUMMATION OF BACKGROUND. A synthesis of perceptual and dialectal variance, this study hopes to push the bounds of perceptual mapping’s known effects. Can perceptual mapping lead to the secondary perception of glides which are fully absent from the sound signal in vowel hiatus sequences? Does this perception differ based on speaker dialect? If phantom glides are perceptible, we expect Appalachian English speakers to perceive them less frequently, as they presumably privilege a vowel mapping to the static, monophthongal form with which they identify. Alternatively, competency with the prestigious, diphthongal form (and its establishment as a perceptually salient, ideal) may affect their preferred mapping (Sumner et al. 2014) and therefore cause the secondary effect of phantom glide perception at an equal level to non-Appalachian speakers. The stigmatization of Appalachian English forms and salient contrast to academic, prestigious, hyper-“correct” forms may affect Appalachian English speakers’ mappings of their own monophthongal vowels.

2. Methods. To determine the limits and robusticity of perceptual mapping’s effects, this study placed Appalachian English’s static, low front monophthongal [aɪ] in hiatus position with [a]. Participants perceived glides in monophthongal [aɪ.a] sequences only if their perception was influenced by perceptual mapping, which incurred the perceived formant transitions necessary to precipitate a phantom glide. To better determine how perceptual mapping might be influenced by a speaker’s native dialect, this study compares Appalachian English speakers with speakers of other American English dialects.

2.1. PARTICIPANTS. Ninety-six native English speakers living within the continental United States participated in the study. Identifying themselves as native Appalachians, 45 participants acquired their first language within the geographic confines of Appalachia and were coded as Appalachian English speakers (mean age = 41, age range = 21-74, 6 cisgender men, 35 cisgender women, 3 of unspecified gender identity). 51 participants did not categorize themselves as native Appalachians and were coded as speakers of other United States dialects (mean age = 30, age range = 19-73, 11 cisgender men, 32 cisgender women, 1 trans man, 2 nonbinary, 1 agender, 4 of unspecified gender identity). Initial participants were recruited through social media advertising (Facebook) and the University of Kentucky's Appalachian Studies listserv. Subsequent Appalachian English participants were recruited through snowball sampling, which entailed asking previous participants to recommend the study to their friends and family members.

2.2. EXPERIMENT DESIGN. Designed as a forced-choice perception task, participants were played a series of stimuli and asked to identify each as one of two options. The stimulus was always an English word suffixed with either [-a] or [-ja]; participants were asked to indicate which suffixed form of each word they heard. As the root words were all vowel-final, the [-a] suffix condition created a hiatus configuration, creating the environment for the perception of acoustic or phantom glides.

The stimuli consisted of three types of words: ten words ending in [i] (type 1), [ə] (type 2), and monophthongal [aɪ] (type 3) were included, giving a total of 30 words. The stimulus types were designed to give rise to the perception of an acoustic glide (type 1), no glide (type 2), or a potential phantom glide (type 3). Stimulus types were designed to create optimal acoustic and phantom glide environments specific to the [-a] condition. The [-ja] suffix condition presented participants with an intentional, phonological glide consonant following word-final vowels. See Figure 1 for a breakdown of comparable populations, stimulus types, and conditions.

Comparable Populations	Stimulus Types	Conditions
Appalachian English Speakers	① i#	-a
		-ja
	② ə#	-a
		-ja
	③ mono. aɪ#	-a
		-ja
Other American English Dialect Speakers	① i#	-a
		-ja
	② ə#	-a
		-ja
	③ mono. aɪ#	-a
		-ja

Figure 1. Experiment conditions breakdown

Type 1 stimuli (i.e. *bee*, *flea*, *knee*) were included to measure the perception of acoustic glides. In the [-a] suffix condition, type 1 stimuli were likely to incur the perception of acoustic glides due to glide-like formant transitions from [i] to the suffixed [a]. Formant transitions from the high, front [i] to the low, back suffixed [a] mimic a phonological glide in the acoustic signal. Type 2 stimuli (i.e. *sofa*, *tuna*, *villa*) were included as controls. When confronted with type 2 stimuli, participants were expected to perceive few glides. The formant transitions from the mid, static [ə] to the low, back suffixed [a] are not drastic enough, and they do not start from a high enough initial articulation to acoustically resemble a glide. Nor did participants have an impetus to map [ə] to a high, front position to incur phantom glide perception. Most directly informative for our research question, type 3 stimuli (i.e. monophthongal *pie*, *spy*, *lie*) were included to test an extreme case of perceptual mapping. A truly low vowel to low vowel sequence, type 3 monophthongal [aɪ] to suffixed [a] stimuli lack the notable F1 formant transition necessary for the perception of an acoustic glide. Therefore, phantom glides perceived from type 3 stimuli suffixed with [a] were the result of perceptual mapping. To perceive a glide, participants mapped a low, front, monophthongal [aɪ] to a high, front position, resembling its diphthongal pronunciation.

The [i] and monophthongal [aɪ]-final stimuli were one syllable English words to avoid stress confounds. The control, type 2, could not be one syllable. Due to English's minimum word requirements, English [ə]-final words must be at least two syllables long. Type 2 stimuli therefore contained an unstressed syllable in between their main stress and the suffix.

A script randomized which condition of each stimulus, either the [-a] or [-ja] suffixed form, was played for each participant.

2.3. STIMULUS CREATION. In total, 60 stimuli were created for this study. All stimuli were recorded by the first author (a 21-year-old native Appalachian English speaker). While producing other stimulus sequences using an approximation of standard American English (*SAE), the first author actively codeswitched to their native Appalachian English to produce stimuli containing monophthongal [aɪ]. For naturalistic formant transitions, each of the 30 English test words was recorded with individual realizations of the [-a] and [-ja] suffixes. The stimuli were recorded in question-answer sentence frames to allow for standard production without prosodic variation (e.g. Which [stimulus] did your sister see? My sister saw the [stimulus] that needed painting). Following recording, stimulus sequences were isolated in Praat (Boersma & Weenink 2001). Spectrograms of examples of each stimulus type in each condition are shown below. Lines delineating sounds are approximate.

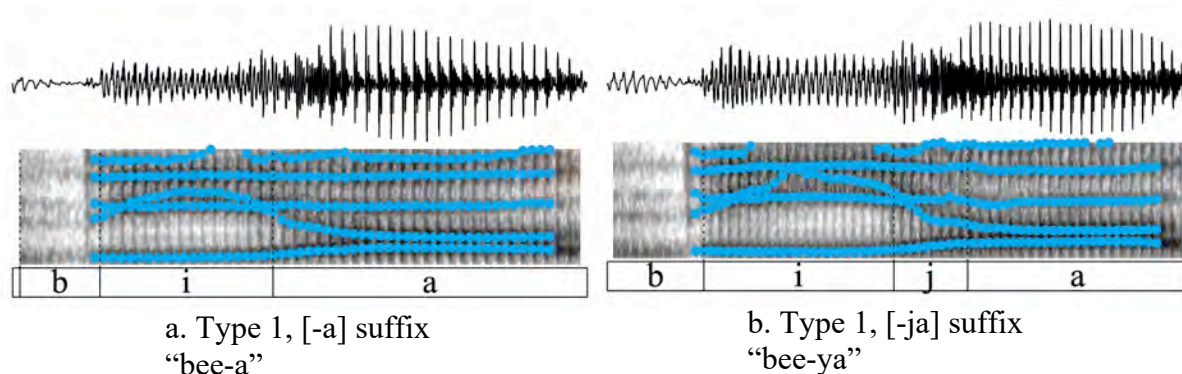


Figure 2.

The spectrograms in Figure 2 exemplify type 1 stimuli. While their formant signatures mimic each other, “bee-ya”, shown in Figure 2(b), was produced with an intentional glide and “bee-a”, shown in Figure 2(a) was produced as a two-vowel sequence. The rise of F1 and fall of F2 necessary to transition from [i] to [a] in Figure (2a) incurs an unintentional acoustic glide. The inherent similarities between the spectrograms reflects the forms’ reasonable confusability.

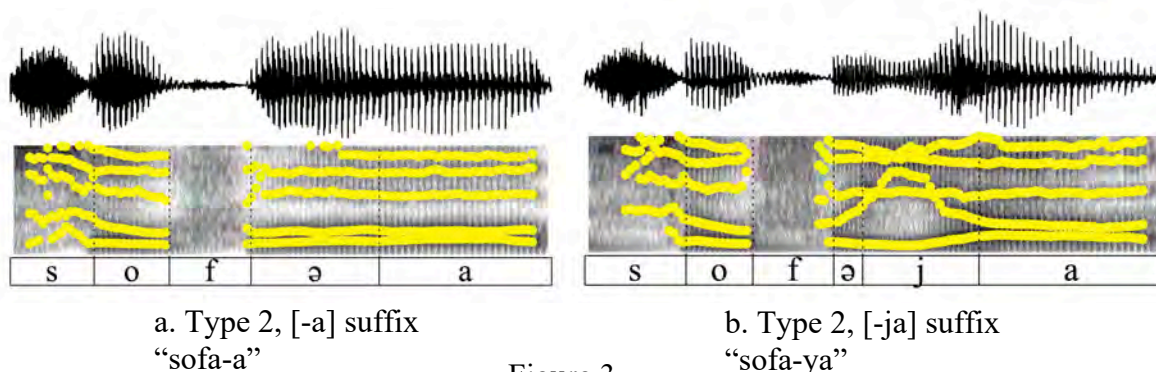


Figure 3.

Examples of type 2, control stimuli Figure 3(a) and Figure 3(b) are distinctly different. Figure 3(a), a type 2 stimulus suffixed with [-a], shows relatively little formant transition between the word-final [ə] in “sofa” and the suffixed [a]. Nothing in the formant signature remotely mimics a glide. Figure 3(b) “sofa-ya”, a type 2 stimulus suffixed with [-ja], is notably different. Following [ə], F1 falls and F2 dramatically plateaus at a high value (indicating a fronter tongue position). F2 proceeds to fall and F1 rises throughout the [j]’s production. As the [j] in Figure 3(b) is phonological and intended, these formant movements are expected.

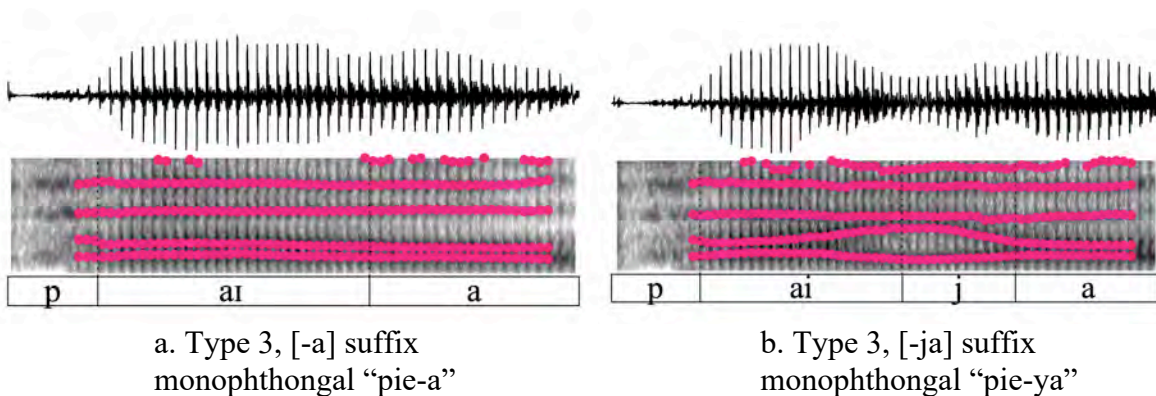


Figure 4.

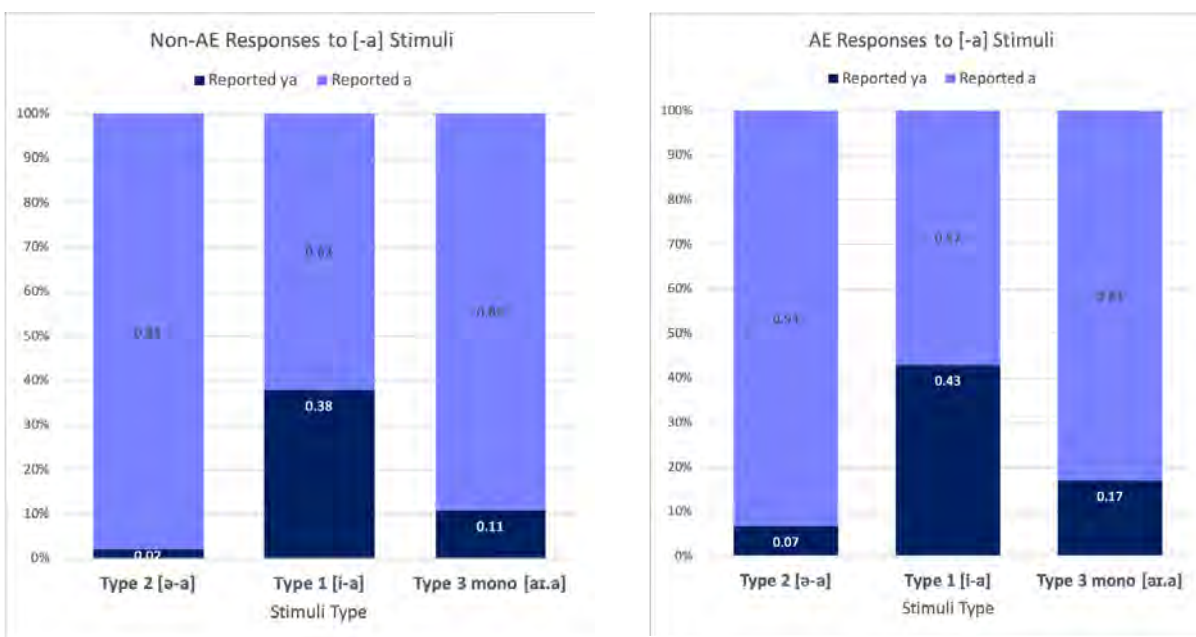
Finally, Figure 4 shows spectrograms of type 3 stimuli. A monophthongal pronunciation of “pie” suffixed with [-a] is shown in Figure 4(a). Note the static formants throughout the monophthongal [ai] production that remain nearly unchanged for a transition to [a]. Nothing in the acoustic signal indicates the presence of a glide. Figure 4(b) shows a monophthongal [ai] suffixed with [-ja]. While the [ai] formants are initially stationary, an unavoidable F1 fall and F2 rise surface as the speaker transitions to the high, front starting position of a phonological glide. The [j] in Figure 4(b) exhibits the archetypical rising F1 and falling F2 of a palatal glide.

2.4. PROCEDURE. This forced-choice task was deployed online through Ibex Farm (Drummond 2016). All participants took part in the study via a weblink without proctoring. They were asked to wear headphones. Following a declaration of consent, participants had the opportunity to play and replay a test sound, which allowed for volume adjustment. Participants were then instructed to assume for the purposes of the study that English has two new suffixes, “-a” and “-ya.” They were told they would be asked to indicate which suffix had been added to 30 different English words. Though encouraged to listen closely, they were asked to go with their first instinct when deciding. Upon hearing each stimulus, two clickable words appeared on the participants’ screens. One clickable word was suffixed with [-a] (e.g. “tree-a”) and the other was suffixed with [-ja] in typical English orthography (e.g. “tree-ya”). Stimuli were not replayable. After making a choice, the screen immediately advanced to the next stimulus, which played without further prompting. The question “Which word did you hear?” remained on the screen for the duration of the study. After identifying their 30th word, participants were presented with a series of demographic questions before exiting the study.

2.5. INCLUSION CRITERIA. Only participants who correctly identified 80% of the suffixes when presented with our control type (type 2) were included in this analysis. Two native Appalachian English speakers were omitted, leaving 43 for analysis. Two speakers of other U.S. dialects were omitted, leaving 49 for analysis.

3. Results.

3.1. RESULTS BY DIALECT. The following figures show responses by participant dialect. The graphs are limited to the [-a] suffix condition, which, in potentially causing the perception of acoustic glides and phantom glides, was our most important stimulus configuration. Responses to [-ya] stimuli can be seen in (3.2) and are discussed in (4.2).



a. Non-AE Responses to [-a] Stimuli
(n=49)

b. AE Responses to [-a] Stimuli
(n=43)

Figure 5.

Figure 5(a) shows the responses of Non-Appalachian English participants to [-a] stimuli. Participants reported hearing glides in 2% of the type 2 (control) sequences. As expected, Non-Appalachian English participants reported hearing the most glides in type 1 (acoustic glide) sequences. Glides were perceived 38% of the time in those [i-a] sequences. Participants perceived glides in type 3 (phantom glide) sequences 11% of the time. Notably more glides were perceived in type 3 (phantom glide) sequences than in type 2 (control) sequences. Incurring the perception of the formant transition necessary to perceive phantom glides, Non-Appalachian English participants sometimes mapped monophthongal [aɪ] as having a high front end point like diphthongal /aɪ/, and thus did in fact hear phantom glides

Figure 5(b) shows the responses of Appalachian English participants to [-a] stimuli. Participants reported hearing glides in 7% of the type 2 (control) sequences. As expected, participants also reported hearing the most glides in type 1 (acoustic glide) sequences. Glides were perceived 43% of the time. Participants perceived glides in type 3 (phantom glide) sequences 17% of the time. Once again, notably more glides were perceived in type 3 (phantom glide) sequences than in type 2 (control) sequences. We see that Appalachian English participants sometimes mapped monophthongal [aɪ] as having a high front end point like diphthongal /aɪ/. This incurred the perception of the vowel distance and transition necessary to perceive phantom glides.

3.2. STATISTICAL MODEL. A generalized linear mixed effects model (GLMM) approach to logistic regression was run in SPSS with *response* as the dependent variable and *participant dialect*, *stimulus type*, and *suffix condition* as fixed effects. The model included random intercepts for the individual participants and for the individual words. Figure 6 details which terms in the model were statistically significant.

Source	F	df1	df2	Sig.
Corrected Model	42.083	11	342	.000
Dialect	1.389	1	151	.240
Type	18.305	2	34	.000
Suffix	413.434	1	2,748	.000
Dialect*Type	1.362	2	2,748	.256
Dialect*Suffix	5.425	1	2,748	.020
Type*Suffix	139.878	2	2,748	.000
Dialect*Type*Suffix	1.048	2	2,748	.351

Probability distribution:Binomial
Link function:Logit

Figure 6. Fixed effects table

As seen above, the three-way interaction between *participant dialect*, *stimulus type*, and *suffix condition* was not significant. The two-way interaction between dialect and type was also

not significant. Together, these results do not support the initial hypothesis that participants respond to our type 3 stimuli differently based on their dialectal affiliation. We conclude there was no difference in the perception of any stimulus type due to participant dialect.

The interaction of *dialect* and *suffix* was statistically significant ($p = .020$). Overall, non-Appalachian English speakers were somewhat more likely than Appalachian English speakers to report hearing “-a” in the [-a] suffix condition and “-ya” in the [-ja] suffix condition. This result does not relate to any aforementioned hypotheses and will not be commented on further.

Supporting basic hypotheses that influenced the experiment’s design, the interaction of *type* and *suffix* proved statistically significant ($p < .001$). The responses to different suffixes varied greatly by type.

Figure 7 graphs the estimated means for the three-way combination of *participant dialect*, *stimulus type*, and *suffix condition*.

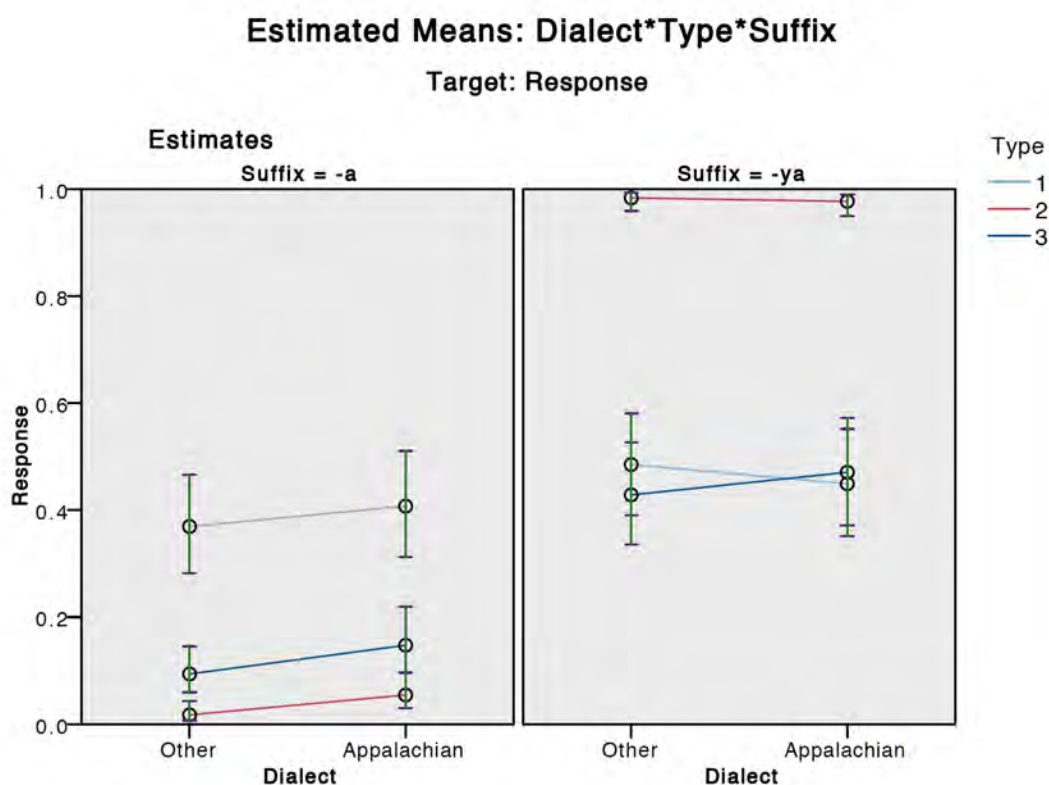


Figure 7. Graph of Estimated Means (three-way interaction)

The estimated probability of participants reporting “-ya” for each type and suffix combination is shown on the y-axis. The graph at left shows probability estimates for the [-a] suffix condition, and the graph at right shows the probability estimates for the [-ja] suffix condition. Stimulus types are differentiated by line color.

Across both graphs, the type 2 (control) condition behaves as expected. Referring to the red line, note the low probability of participants responding “-ya” when played a type 2 stimulus suffixed with [-a] (at left). Note the exact, expected reversal of type 2 responses in the graph at right. Participants are expected to reliably respond “-ya” when played type 2 stimuli suffixed with [-ja]. As the responses to [-ja] suffixes are not informative of the perception of acoustic or phantom glides, I will now focus exclusively on the “-a” suffix estimated means (graph at left).

Estimated type 1 stimulus responses, shown with a light blue line, appear to indicate confusion attributed to the perception of acoustic glides. If played a type 1 stimulus with an [-a] suffix, the model predicts participants will categorize it as “-ya” around 40% of the time. The type 3 estimated responses, which are indicated in dark blue, show less expected “-ya” responses than type 1 stimuli but notably more “-ya” responses than type 2 stimuli. The dark blue line is statistically significantly higher ($p < 0.001$; pairwise comparison) than the red line (which refers to responses from the type 2, control condition) showing that phantom glides were in fact perceived. The lack of significant three-way interaction term shows us that phantom glide perception occurred across participant dialect groups.

Pairwise contrasts comparing responses of each stimulus type within each suffix condition are shown in Figure 8.

Suffix	Type Pairwise Contrasts	t	df	Significance
-a	1 - 2	± 9.513	185	0.000
	1 - 3	± 7.081	46	0.000
	3 - 2	± 4.365	221	0.000
-ya	1 - 2	± 11.806	264	0.000
	1 - 3	± 0.370	27	0.715
	2 - 1	± 11.990	266	0.000

Figure 8. Pairwise contrasts

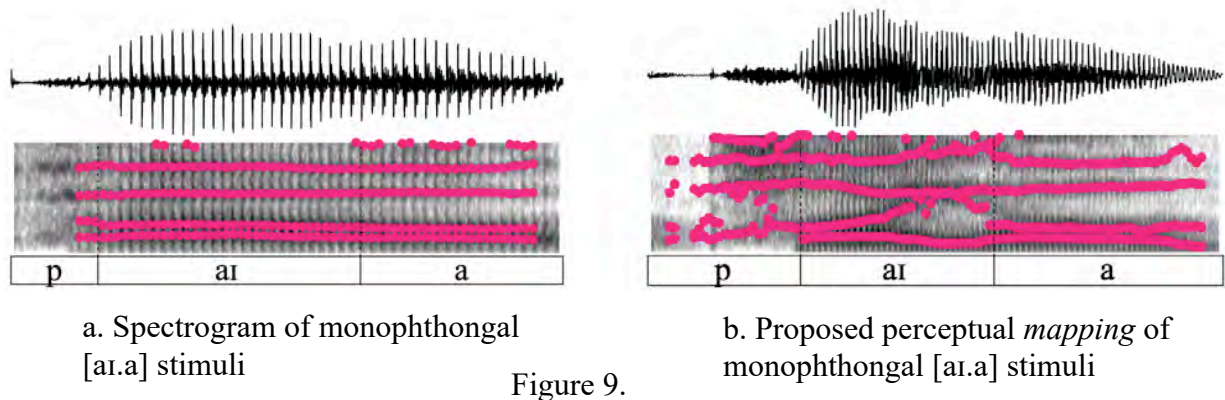
Responses to the [-a] suffix are significantly different across all three stimulus types. The stimulus types, which incur the perception of nearly no glides (type 2, control), acoustic glides (type 1), and phantom glides (type 3), have statistically significantly different responses. Perhaps surprisingly, there is no contrast between type 1 and type 3 stimuli within the [-ja] suffix condition. This deviation will be discussed in section (4.3).

4. Discussion.

4.1. ACOUSTIC GLIDES. As expected by Hogoboom’s (2020) findings, acoustic glides were confusable with phonological glides for all participants. Cross-dialectally, type 2 (acoustic glide) stimuli suffixed with [-a] were perceived as suffixed with [-ja] in 40% of instances. Similarly, type 2 stimuli suffixed with a phonological glide, [-ja], were perceived as suffixed with [-a] in 53% of instances. Perception of a glide in an [i.a] sequence appeared close to chance; participants were largely unable to perceive a difference between type 2 phonological glide ([-ja] suffixed) and acoustic glide ([-a] suffixed) stimuli. The rise of F1 and fall of F2 necessary to transition from [i] to [a] in hiatus sequences too closely mimics the rise of F1 and fall of F2 in the articulation of a phonological [j] for our participants to differentiate the intended form.

4.2. THE EXISTENCE OF PHANTOM GLIDES. Averaged across all participants, type 3 (phantom glide) stimuli suffixed with [-a] were perceived as suffixed with [-ja] in 14% of instances. Participants in both dialect groups perceived a glide that was truly absent from the speech stream. To

perceive phantom glides, participants from both dialects at times mapped static formants in monophthongal [aɪ.a] stimuli (Figure 9a) to a diphthongal [aɪ] with a high front endpoint (Figure 9b), which incurred the necessary F1 fall and rise for glide perception.



Given that speakers can map monophthongal [aɪ] to diphthongal /aɪ/, we might wonder why this does not happen more frequently. The lack of more robust phantom glide perception may support the conception of a dual-representation model, wherein both Appalachian English and Non-Appalachian English speakers have underlying forms for monophthongal and diphthongal /aɪ/. Using exemplar-based logic, Non-Appalachian speakers have access to monophthongal /aɪ/ exemplars through media, cross-dialectal contact, and, as stated, may subconsciously use /aɪ/ production as a heuristic for identifying Southern American English speakers (Clopper & Pisoni 2004, 2007). Similarly, though of a distinct power differential, Appalachian English speakers have access to diphthongal /aɪ/ exemplars through educational settings, mass media, and cross-dialectal contact.

We therefore assume both populations have access to both the underlying representation of monophthongal and diphthongal /aɪ/. This means populations can correctly map the monophthongal [aɪ]. However, participants sometimes mapped it to an erroneous high front position, incurring the perception of phantom glides. This mapping outweighed phonetic detail from a continuous speech stream. We propose the diphthongal underlying representation has a prestige pull in the perceptual system through *social-weighting* (Sumner et al. 2014). Their model uses social-weighting to explain memory inequality and unequal priming effects. We posit it lures participants to make erroneous mappings in favor of stronger, socially weighted forms. Acoustic patterns and social representations are mapped in tandem via a dual-route in their expanded exemplar model. Following this dual-route approach to speech perception, diphthongal /aɪ/ percepts would retain higher degrees of saliency in the lexical access and speech processing system due to the qualitative social information stored about the form.

How can such social weighting lead to Appalachian English speakers' privileging of an underlying form with which they do not identify? While privilege may be present across broader linguistic contexts due to social-weighting in the speech processing system, the academic register of this study may have also encouraged Appalachian English speakers to invoke their most prescriptive language forms. Rather than (or perhaps in conjunction with) subconscious, processing bias, Appalachian English participants may have consciously accessed what they assumed to be their most "correct" underlying forms.¹ We cannot disregard the potential for unintended effects

¹ Anecdotally, multiple Appalachian English participants expressed hope that they had "gotten a good score" after completing the study or expressed misgivings and worry that their answers were "wrong."

wrought by this experiment's procedure and presentation. For instance, choosing between options like "pie-a" and "pie-ya" likely induced reading-specific effects. Ziegler, Muneauz, and Grainger (2003) have suggested that there are orthographic effects on speech processing; an orthographic form can restructure phonological representation. The prescriptivist conventions learned by English readers and writers likely shape their processing of spoken language. Meta-linguistic commentary about standards might encourage the usage of exemplars built from experiences of careful, hyper-articulation to be used in formal disambiguation contexts (Sumner et al. 2014).

4.3. THE [-JA] SUFFIX CONDITIONS. Type 2 (control) stimuli in the [-ja] suffix condition behaved as expected. Participants overwhelmingly correctly identified [ə-ja] stimuli as a "-ya" from. While we did not have a clear hypothesis for their treatment by participants, type 3 stimuli in the [-ja] suffix condition were perceived similarly to type 1 stimuli. They were both very frequently reported as being affixed with the [-a] suffix

As shown in Figures (2a) and (2b), type 1 stimuli in the [-a] and [-ja] conditions were acoustically similar. Perhaps understanding the inherent formant change necessary in type 1 sequences, participants "factored out" an intentional [j], attributing it to mere vowel transitions.

Our type 3 stimuli, shown in Figure 4, were radically different across suffix conditions. Form confusability cannot be used to explain [-a] responses in type 3 [-ja] conditions. Instead, we propose that participants mapped the beginning of the suffixed [-ja] in our type 3 stimuli, which has an unavoidable F1 fall, as part of the preceding vowel. Essentially, the transition from a monophthongal [aɪ] to the high front starting place of an intentional [j] was perceived as movement to a secondary vowel target region and not the beginning of a consonant; monophthongal [aɪ-ja] was sometimes perceived as diphthongal [aɪ-a]. This result can also be explained by the social weighting of the diphthongal /aɪ/ form.

5. Conclusion. Participants across dialect groups perceived a consonant that was truly absent from the speech stream. Processes of perceptual mapping in our type 3 [-a] suffix condition eclipsed cues from the stimuli's continuous acoustics, incurring the perception of phantom glides. As speakers did not hear phantom glides with overwhelming frequency, we reasonably assume they otherwise mapped monophthongal [aɪ] to a low front underlying form. The privileging of an unfounded diphthongal /aɪ/ mapping instead of a low front mapping that participants appeared to possess suggested an imbalance in the perceptual saliency of the forms. Diphthongal /aɪ/ is marked as overtly more prestigious through meta-linguistic commentary, academic instantiation, and mass media. Storing such qualitative components of the form through socially weighted encoding (Sumner et. al), listeners may develop conscious and subconscious perceptual bias for diphthongal /aɪ/ and other prestigious forms.

Because it is necessary for participants recognize the English word being affixed, our study used standard orthography of the words for the response choices. In a follow up study, we will use similar stimuli (types 1, 2, and 3 across both suffix conditions), but we will embed the test words in spoken sentences in order to circumvent an academic, orthographical mapping bias.² Participants will still be asked to indicate the test word's perceived suffix, but they will select from standard "-a" or "-ya" options. Because the word will be recognizable from the context of the sentence, the words will not need to be presented orthographically. Additionally, we plan to

² Many thanks to Paul Reed and Nathan Sanders for their help conceptualizing this follow-up.

use these larger sentences as an opportunity for priming. The full sentence will be recorded in an Appalachian English dialect. Such attempted priming will help us better understand the robusticity of a perceptually privileged diphthongal /ai/. In a more vernacular, Appalachian English context, Appalachian English speakers (and perhaps speakers of other American dialects) may map monophthongal [ai] to a low front underlying form, perceiving fewer phantom glides. Conversely, if phantom glides are perceived in Appalachian English contexts, the social-weighting of diphthongal /ai/ may be thought to fundamentally shape participants' perceptual systems.

References

- Bernstein, Cynthia. 2006. Drawing out the /ai/. In Thomas E. Murray & Beth Lee Simon (eds.), *Language variation and change in the Midland: A new look at 'heartland' English*, 209–232. Amsterdam: John Benjamins.
- Boersma, Paul & David Weenink. 2019. Praat: Doing phonetics by computer [Computer program]. Version 6.1.05. Retrieved from <http://www.praat.org/>.
- Burgdorf, Dan C. & Sam Tilsen. 2021. Glides prioritize articulation, vowels prioritize acoustics. Talk given at the 2021 Annual Meeting of the Linguistic Society of America, January 7–10.
- Casali, Roderic F. 2011. Hiatus resolution. In Marc van Oostendorp, Colin J. Ewen, Elixabeth Hume & Keren Rice (eds.), *The Blackwell companion to phonology*. Online: Wiley. <https://doi.org/10.1002/9781444335262.wbctp0061>.
- Clopper, Cynthia G., and David B. Pisoni. 2004. Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics* 32(1). 111–140. [https://doi.org/10.1016/s0095-4470\(03\)00009-3](https://doi.org/10.1016/s0095-4470(03)00009-3).
- Clopper, Cynthia G. & David B. Pisoni. 2007. Free classification of regional dialects of American English. *Journal of Phonetics* 35(3). 421–438. <https://doi.org/10.1016/j.wocn.2006.06.001>.
- Drummond, Alex. 2016. Ibexfarm. <https://github.com/addrummond/ibexfarm>.
- Fridland, Valerie & Tyler Kendall. 2011. Exploring the relationship between production and perception in the mid front vowels of U.S. English. *Lingua* 122. 779–793. <https://doi.org/10.1016/j.lingua.2011.12.007>.
- Goldinger, Stephen. 1996. Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22(5). 1166–1183. <https://doi.org/10.1037/0278-7393.22.5.1166>.
- Greene, Rebecca. 2010. *Language, ideology, and identity in rural eastern Kentucky*. Stanford, CA: Stanford University dissertation. <https://purl.stanford.edu/fh361zh5489>.
- Hazen, Kirk & Sarah Hamilton. 2008. A dialect turned inside out: Migration and the Appalachian diaspora. *Journal of English Linguistics* 36. 105–128. <https://doi.org/10.1177/0075424208317127>.
- Hazen, Kirk. 2020. *Appalachian Englishes in the twenty-first century*. Morgantown: West Virginia University Press.
- Hogoboom, Anya. 2020. Phonological vs. acoustic glides. Poster given at the Berkeley Linguistics Society Workshop on Phonological Representations. University of California, Berkeley, February 7–8.
- Johnson, Keith. 1997. Speech perception without speaker normalization: An exemplar model. In Keith Johnson, & J. Mullennix (eds.), *Talker variability in speech processing*, 145–165. San Diego: Academic Press.
- Johnson, Keith. 2005. Decisions and mechanisms in exemplar-based phonology. *UC Berkeley PhonLab Annual Report* 1. 289–311. <https://escholarship.org/uc/item/8kf3z3s9>.

- Labov, William, Sharon Ash & Charles Boberg. 2005. *The atlas of North American English: Phonetics, phonology and sound change*. Berlin: De Gruyter Mouton.
<https://doi.org/10.1515/9783110167467>.
- McLennan, Conor, T. Paul Luce & Jan Charles-Luce. (2003). Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29. 529–553.
<https://doi.org/10.1037/0278-7393.29.4.539>.
- Niedzielski, Nancy & Dennis R. Preston. 2003. *Folk linguistics*. Berlin: De Gruyter Mouton.
- Padgett, Jaye. 2008. Glides, vowels, and features. *Lingua* 118(12). 1937–1955.
<https://doi.org/10.1016/j.lingua.2007.10.002>.
- Preston, Dennis R. 1989. *Perceptual dialectology*. Dordrecht: Foris.
- Preston, Dennis R. 1993. Folk dialectology. In Dennis R. Preston (ed.), *American dialect research: Celebrating the 100th anniversary of the American Dialect Society*, 333–378. Amsterdam: John Benjamins.
- Ranbom, Larissa & Cynthia Connine. 2007. Lexical representation of phonological variation. *Journal of Memory and Language* 57. 273–298. <https://doi.org/10.1016/j.jml.2007.04.001>.
- Reed, Paul E. 2014. Inter-and intra-generational /ai/monophthongization, indexicality, and Southern Appalachian identity. *Southern Journal of Linguistics* 38(1). 159–194.
- Reed, Paul E. 2016. *Sounding Appalachian: /ai/ monophthongization, rising pitch accents, and rootedness*. Columbia, SC: University of South Carolina dissertation.
<https://scholarcommons.sc.edu/etd/3563>.
- Reed, Paul E. 2018. The importance of Appalachian identity: A case study in rootedness. *American Speech: A Quarterly of Linguistic Usage* 93(3-4). 409–424.
- Samuel, Arthur & Saioa Larraza. 2015. Does listening to non-native speech impair speech perception? *Journal of Memory and Language* 81. 51–71.
<https://doi.org/10.1016/j.jml.2015.01.003>.
- Sumner, Meghan & Arthur G. Samuel. 2009. The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language* 60(4). 487–501.
<https://doi.org/10.1016/j.jml.2009.01.001>.
- Sumner, Meghan & Reiko Kataoka. 2013. Effects of phonetically-cued talker variation on semantic encoding. *The Journal of the Acoustical Society of America* 134(6). EL485–EL491.
<https://doi.org/10.1121/1.4826151>.
- Sumner, Meghan, Seung Kyung Kim, Ed King & Kevin B. McGowan. 2014. The socially weighted encoding of spoken words: A dual-route approach to speech perception. *Frontiers in Psychology* 4. 1015. <https://doi.org/10.3389/fpsyg.2013.01015>.
- Thomas, Erik. 2001. *An acoustic analysis of vowel variation in New World English* (Publication of the American Dialect Society 85). Durham, NC: Duke University Press.
- Uffman, Christian. 2007. Intrusive [r] and optimal epenthetic consonants. *Language Sciences* 29(1). 451–476. <https://doi.org/10.1016/j.langsci.2006.12.017>.
- Williams, John Alexander. 2002. *Appalachia: A history*. Chapel Hill, NC: University of North Carolina Press.
- Wolfram, Walt. 1977. On the linguistic study of Appalachian speech. *Appalachian Journal* 5(1). 92–102.
- Ziegler, Johannes C., Mathilde Muneaux & Jonathan Grainger. 2003. Neighborhood effects in auditory word recognition: Phonological competition and orthographic facilitation. *Journal of Memory and Language* 48. 779–793. [https://doi.org/10.1016/S0749-596X\(03\)00006-8](https://doi.org/10.1016/S0749-596X(03)00006-8).